DOCUMENT RESUME

ED 085 682 CS 000 881

AUTHOR Corman, Louise; And Others
TITLE More Effective Reading Tests of Comprehension and Rate.
INSTITUTION Research Inst. for Educational Problems, Cambridge, Mass.
SPONS AGENCY Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C.; National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 73
GRANT NE-G-00-3-0016; OEG-0-8-080506-4597
NOTE 29p.; Studies in Learning Potential, Volume 3, Number 65

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Elementary Education; Grade 5; Grade 7; Item Analysis; *Norm Referenced Tests; *Reading Comprehension; Reading Materials; Reading Skills; *Reading Speed; *Reading Tests; Standardized Tests; *Test Construction; Test Validity

ABSTRACT

Reading specialists have often expressed dissatisfaction with standardized reading tests, especially tests of rate. In this study, norm-referenced survey tests were constructed for measuring reading comprehension and rate of fifth and seventh grade students. Three alternate forms were developed from item analyses with a trial sample of 88 fifth grade and 95 seventh grade students in a Boston suburb. The revised forms were then administered to a new sample of 159 fifth grade and 157 seventh grade students from the same school system. The three test forms for each grade showed evidence of internal consistency, validity, a broad possible score range, and equivalence with the other forms. (Author)

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 085682

CS

STUDIES IN LEARNING POTENTIAL

MORE EFFECTIVE READING TESTS OF COMPREHENSION AND RATE

Louise Corman, Joel Weinberg, and Milton Budoff

Volume 3, Number 65

1973

Research Institute For Educational Problems
12 Maple Avenue Cambridge, Massachusetts

FILMED FROM BEST AVAILABLE COPY

AS 000 881

MORE EFFECTIVE READING TESTS OF COMPREHENSION AND RATE

Abstract

Reading specialists have often expressed dissatisfaction with standardized reading tests, especially tests of rate. In this study, norm-referenced survey tests were constructed for measuring reading comprehension and rate of fifth and seventh grade students. Three alternate forms were developed from item analyses with a trial sample of 88 fifth grade and 95 seventh grade students in a Boston suburb. The revised forms were then administered to a new sample of 159 fifth grade and 157 seventh grade students from the same school system. The three test forms for each grade showed evidence of internal consistency, validity, a broad possible score range, and equivalence with the other forms.

MORE EFFECTIVE READING TESTS OF COMPREHENSION AND RATE [1]

Louise Corman, Joel Weinberg, and Milton Budoff
Research Institute for Educational Problems

Several standardized reading tests are currently available which provide researchers and school personnel with quantitative measures of comprehension and rate. Despite the proliferation of these instruments, reading specialists have frequently expressed dissatisfaction with procedures commonly used in constructing reading tests (Anderson, 1972; Chall, 1967). The goal of this study was to construct instruments for measuring reading comprehension and rate of students in the intermediate and junior high school grades. In an effort to overcome some of the frequent objections that have been raised toward methodological procedures, tests were developed in accordance with measurement theory and findings of prior research in reading.

The inclusion of measures of comprehension and rate was considered essential in the tests because adequate facility in both areas is required of students in these grades. By the fourth grade, the student with normal reading development is expected to have mastered word recognition skills and to be capable of processing several words at a time. Word-by-word reading, common in the primary grades, is gradually replaced by a more rapid rate

which accompanies the ability to grasp meaning from larger thought units (Huey, 1968). The rapid reader is unhampered by having to focus on short units of recognition and is therefore free to devote himself to thought interpretation (Buswell, 1922; Judd, 1918). Comprehension and rate become integrally related aspects of the reading process.

The developer of a reading test is posed with a problem which stems from the lack of consensus among reading specialists concerning the definition of comprehension. Approaches toward a definition include skill lists and taxonomies, as well as factor analytic and correlational analyses. Lists of comprehension skills presented by various authors usually contain many of the same skills (Harris, 1961; Smith & Dechant, 1961). The fact that no two authors' lists are identical, however, suggests the difficulty in defining comprehension. Factor analytic studies have indicated that skills enumerated in skill lists and taxonomies are not independent. Rather, studies by Davis (1944, 1971) have revealed two major factors in comprehension: word knowledge and verbal reasoning. Traxler (1958) held that the verbal reasoning factor identified by Davis is very similar to general intelligence as measured by tests of mental ability. He obtained correlations from .65 to .74 between reading scores on three standardized tests and Kuhlmann-Anderson mental age

scores with second through eighth grade students (Traxler, 1941a).

Unlike comprehension, reading rate is relatively simple to define and can be more easily measured. Rate can be defined as the rate at which material is read with some degree of comprehension. Few standardized tests for young children include a measure of rate. Those that do, measure comprehension and rate simultaneously and are usually presented in a format of several short passages, with multiple-choice questions following each passage. Scores on several comprehension subtests may be obtained. Time limits are often imposed, and rate may be calculated as the number of exercises attempted in a given time or as the number of questions answered correctly within the time limit.

The reading literature abounds with criticisms of these tests, and they are often distrusted by teachers and principals (Chall, 1967). The following objections are frequently cited:

1. The fact that many comprehension tests are timed decreases the validity of those tests as measures of comprehension, because comprehension becomes confounded with reading rate (Chall, 1958; Harris, 1961; Smith & Dechant, 1961). Furthermore, reliability of timed rate tests is often spuriously inflated by the speed element (Spache, 1963).

2. Tests which insert comprehension questions into

the reading text penalize the slow reader on his comprehension score (Spache, 1963; Traxler, 1941b).

3. Stability of rate scores obtained over a very short time interval is doubtful (Traxler, 1958). Tests which score rate as the number of questions answered correctly in a given time are confounding comprehension with rate and therefore do not give a pure measure of either ability (Harris, 1961). On the other hand, rate scores based on the number of questions attempted in a given time are not accurate because some difficult questions may be tried but left unmarked (Robinson & McCollum, 1934).

4. There is rarely enough time in a class period to include a large number of items in any one subtest. As a result, subtests designed to measure separate elements of comprehension are rarely reliable (Smith & Dechant, 1961; Traxler, 1958).

5. Standardized tests which use selections for typical students in a range of grade levels may give a distorted picture of the true achievement level of students who are retarded or advanced for their grade (Chall, 1958).

6. Certain tests are too time consuming to administer and score (Harris, 1961; Traxler, 1958).

7. Many questions on comprehension tests can be answered without reading the passages, casting doubt on the validity of such tests (Simons, 1971).

In this study three alternate forms of norm-referenced

survey tests for measuring reading comprehension and rate at each grade level (5 and 7) were constructed. Four forms were initially constructed for each grade to permit selection of the three forms which best met statistical criteria of norm-referenced tests. Each of the four forms contained a fiction and a factual selection.

The following procedures were adhered to in order to avoid the limitations previously cited:

1. The tests are not timed and are constructed so that nearly every student can complete reading the selection and answering the questions.

2. Each reading selection is 1100 to 1500 words in length, with questions following, rather than inserted into the text.

3. The rate score is the mean number of words per minute read over three consecutive one-minute time periods, the minimum interval considered necessary for a reliable rate measure (Traxler, 1958).

4. Four comprehension skills are included but are not considered as separate subtests, since few items on each skill can be given in one class period. Only total comprehension scores are used.

5. The selections and questions on any one test are geared to students of a particular grade level, and the tests allow for a broad range of possible scores for students in that grade.

6. Each form can be administered in a forty-five-minute class period and can be scored by computer.

7. The comprehension questions alone were given to a comparable group of fifth grade students, to determine whether the questions could be answered without prior reading of the selections.

Test development was conducted in three phases: (a) construction of original forms, (b) test development, i.e., revision of items based on item analysis with a sample population, and (c) evaluation of the psychometric characteristics of final test form..

Construction of Original Test Forms

Comprehension items for each form were developed in accordance with the test construction procedures recommended by Furst (1958) and Bloom, Hastings, and Madaus (1971). Initially, a table of specifications (Table 1) was developed, which delineated the content and behaviors selected for measurement. Given the inconsistency among lists of skills thought to comprise comprehension, as well as research results indicating the overlap of those skills, the authors did not seek to define the universe of skills which might constitute comprehension. Rather, a small number of skills (recall of main ideas,[2] recall of minor details, recall of sequences, and ability to draw inferences) were selected as indicators of comprehension. These skills have been stressed

in the literature and the extent of their use in the classroom has been clearly documented (Guszak, 1967).

Insert Table 1 about here

Two reading specialists wrote items to fill each of the eight cells in the table. Each specialist was asked to indicate which behavior in the table corresponded to each item on each selection. The degree of agreement between the two specialists attested to the content validity of the test (Bloom et al., 1971).

The number of items in each cell in the original test forms varied. It was considered desirable to try out several formats for certain types of questions in order to permit selection of the format which best met statistical criteria for the final forms. Since the particular reading selection somewhat determined the number of questions that could be asked for each behavior, the total number of items and the number in each cell differed from one form to another. On these original forms, there was a minimum of 18 questions on each selection and 36 questions on each total form.

The format of each question was multiple choice with four options. Items on the original forms were, in general, ordered according to the order in which they appeared in the story. The distribution of correct responses was determined through the use of a table of random digits, as

TABLE 1

Table of Specifications for Comprehension Questions according to Barrett's "Taxonomy of Cognitive and Affective Dimensions of Reading Comprehension (Clymer, 1968)"

	Behaviors			
Content	1.21 Recall of details	1.22 Recall of main ideas	1.23 Recall of sequences	3.0 Inferential comprehension
Fiction	4	4	2	4
Fact	4	4	2	4

Note.--The number of items in each cell pertains to the final 28-item forms. Figures for original forms varied.

a check against the influence of response sets (Furst, 1958). A computer scoring program was written to insure scoring objectivity.

Because fiction is assumed to be generally easier to read than nonfiction, the fiction selection and the questions following it preceded the factual selection and its questions in each test booklet. (This assumption was tested in the test development phase.) This procedure was an attempt to reduce the anxiety of the examinee.

Directions to examinees were intended to be as free from ambiguity as possible. The examiner distributes the tests face down and says:

When I ask you to, turn the booklet over and read the first story straight through for good comprehension. While you are reading the story, circle the word you are reading each time you are instructed to do so. When you finish reading the story, answer the questions that follow the story. Do not look back at the story while you are answering the questions. When you finish answering the questions, turn the booklet over.

The tests are not timed. When all students have turned over their booklets, or when 20 minutes have elapsed, the examiner repeats these instructions for the second story. The 20 minute period was found to be long enough for nearly every student to complete the reading and test items.

At precisely the end of the first, second, and third minutes during the reading of each selection, the examiner

asks the students to circle the word they are reading. Three rate scores are calculated: the mean number of words read on the three one-minute intervals for the fiction and the factual selections, and the six-minute mean for the total test. Three comprehension scores can be computed: the percentage of correct responses of the total number of questions on the fiction selection, the factual selection, and the total test form.

Directions for administration indicate that the students are not instructed to read as fast as possible, a procedure used in some experiments to stimulate a purpose for rapid reading. It was felt that such directions could produce rates that might not reflect the pupils' normal reading speed and could distort comprehension scores at the same time. Students are instructed not to look back at the story while answering questions to insure that recall of information is measured.

The fact that no instructions are given concerning the advisability of guessing leaves the decision of whether to guess up to the student. This procedure does not eliminate individual differences in gambling tendencies but attempts to control their effects on scores. If item difficulties are appropriate, guessing by some individuals is to some extent accounted for (Furst, 1958).

Test Development and Revision

The following questions were answered during this phase

of the study:

1. Which items on each form should be selected for inclusion in the final 28-item forms?

2. Does the average item difficulty for each selection of each form allow for a broad range of total scores among students in the grade for which that form was designed?

3. Can the three forms for each grade level be considered statistically equivalent?

4. Does the order in which a student receives the test forms affect his performance in rate of comprehension?

5. Is there a practice effect on rate or comprehension over repeated testing?

6. Is there a difference between rate or comprehension scores of the same subjects on fiction and factual materials on any form?

7. Can a comparable group of fifth grade subjects attain similar mean comprehension scores on either selection of any fifth grade form, without having read the selection?

Subjects

The sample consisted of 88 fifth grade and 95 seventh grade students in a suburb of Boston, who were approximately evenly divided into four classrooms at each grade level. The fifth grade sample was drawn from two schools and the seventh graders from one junior high school. The population in the school districts from which the sample was drawn consists mainly of working class and middle class families.

A diversity of ethnic backgrounds was represented, including many students of Armenian, Italian, and Irish descent. The majority of the sample was Catholic or Armenian Orthodox. No blacks and few Jews were included. Since this school system offers special classes for pupils with IQs below 80, no students with extremely low IQs were included in the study.

Procedure

During the course of a one-month period, each student was tested in four 45-minute class periods, taking one of the four test forms constructed for his grade level at each test administration. Each classroom at each grade level was randomly assigned to one of four sequence groups, which determined the order in which the students took the four test forms. To control for carry-over effects (Winer, 1962), the four orders represented by the sequence groups were chosen from all possible orders in accordance with a balanced Latin square design. At the end of the one-month period, each student took the Reading Subtest of the Metropolitan Achievement Tests; fifth grade students received Intermediate Form A and seventh graders Advanced Form A.

After all tests had been scored, item analyses on each selection were performed for the purpose of revising the original tests. These item analyses were based on item scores of the total sample receiving a given form, without regard to the particular test session on which that form was administered.

According to the difficulty and discrimination indices of each item, fourteen items were selected for retention in a final version of each selection. The total test therefore consisted of 28 items: each of the two selections had four items about main ideas, four about minor details, four on inference, and two on sequence (see Table 1). Only two sequence questions were included, because the content of a selection restricted the number of possible sequence questions to a greater extent than it did the other three item types. The final test forms, then, had a uniform number of total items and items within each type.

Items were selected which had the highest discrimination indices and difficulty levels closest to .50. This procedure was used to increase the reliability of each form and to allow for a broad possible range of scores, reducing the likelihood of a ceiling effect. When there were too few items of a particular type which met these criteria, existing items were rewritten to make them easier if their difficulty level was under .20, or harder if their difficulty level was over .80. If it appeared that rewriting an existing item might not correct it enough to meet the statistical criteria (e.g., when the discrimination index was negative), a new item was constructed in order to have the required number of items of each item type.

In order to have a parallel pattern for the final

test forms, items on each selection were repositioned according to their type. To lessen the students' anxiety, the first two items on each selection were the easiest of their type, as indicated by the difficulty levels obtained.

After all tests had been revised, the original tests were rescored, using only those items which had been selected for retention in the final forms. These revised scores were used in all statistical analyses in this phase of the study. The assumption was made that repositioning and deletion of some items would not affect the statistical properties of the remaining items when the final forms were used.

Results

Fifth grade forms will be referred to as 5A, 5B, 5C, and 5D, and seventh grade forms as 7A, 7B, 7C and 7D.

Mean comprehension scores, reflecting the average difficulty level of items selected for inclusion in the revised test forms ranged from 50 to 66 percent. The absence of a ceiling effect was considered favorable for allowing a broad range of scores. Mean rates were similar on forms at each grade level.

Evidence of equivalence among the four forms at each grade level was provided by the similarity of means and standard deviations on all test forms except Form 5C. This form was found to produce large variance in scores that were

influenced by an order of administration effect. Additional evidence of form equivalence was given by alternate form correlation coefficients which ranged from .60 to .72 in all cases except those involving Form 7A.

The internal consistency of all forms was demonstrated by KR20 coefficients on comprehension that ranged from .74 to .89. These coefficients confirmed that the difficulty levels and discrimination indices of items selected for the revised forms were appropriate, and that the items selected were contributing to the homogeneity of the test forms.

Evidence of concurrent validity was provided by correlations between comprehension scores on each form and scores on the Metropolitan Reading Test which ranged from .58 to .73. Validity pertaining to the internal homogeneity of these tests, which were designed to measure one construct, was attested to by the KR20 coefficients (Cronbach & Meehl, 1967). Content validity was insured during construction of the original tests when two reading specialists selected items for each cell in the table of specifications. Details of statistical characteristics of each form are presented in Corman (1973).

To test the effects of order of administration, test form, and test session, a Latin square within a repeated measures design was used. Two repeated measures analyses of variance per grade level were performed on rate and comprehension. Test Form and Test Session were within

subjects factors and Sequence Group was the among subjects factor. Results indicated that random assignment of intact classrooms to sequences of test forms for the four test administrations failed to prevent a significant effect due to the order in which a student received a particular form. Significant interactions between Test Form and Test Session on comprehension at both grade levels ($p < .05$) and on rate for seventh graders ($p < .001$) showed that the order in which a student received a form influenced his scores. The decision was made, therefore, to assign individuals randomly to forms when characteristics of the final test forms were evaluated.

A significant linear relationship between Test Session and rate ($p < .001$) indicated that a practice effect over the four test administrations had occurred with reading rate. Students at both grade levels read faster at each successive test administration. Repeated testing did not result in a significant practice effect on comprehension scores.

The effect of content on rate and comprehension was found through analyses of variance to vary with different test forms. For those forms on which scores were differentiated by type of content, fiction appeared to be generally easier than fact, in terms of both rate and comprehension.

In order to determine whether subjects could guess the correct answers to the comprehension questions without

having read the selections, all fifth grade forms were given to a new sample of 17 fifth graders who attended one of the two elementary schools from which the target population was drawn. The four forms were distributed randomly. (Unfortunately, it was not possible to obtain a seventh grade sample for this comparison.) One way analyses of variance indicated that scores of fifth graders who answered the comprehension questions without prior reading of the material were significantly lower than those of students who had read the selections, on every form except 5C ($p < .01$).

With the exceptions of Forms 5C and 7A, all forms were considered to possess an acceptable degree of reliability, validity, and equivalence with the other forms of the same grade level. These two forms were then eliminated from use in the final test battery. The remaining six forms were renamed 05, 06, 08, 10, 11, and 12, respectively, in order to distinguish then from the original test forms.

Characteristics of Final Test Forms

The purpose of this phase of the study was to examine psychometric characteristics of the six test forms developed after items on each original form had been selected and revised. Determination of reliability and concurrent

validity of the final forms was considered especially important in this phase.

Subjects

The sample for this phase of the study consisted of 159 fifth grade and 157 seventh grade students, divided fairly evenly into six fifth grade classrooms (two in each of three schools) and five seventh grade classrooms from one junior high school. Students resided in the same community from which the test development sample was drawn; however, no schools participating in the test development phase were included. Characteristics of this sample were similar to those of the sample previously described.

Procedure

Each student, regardless of his grade placement, took a fifth and a seventh grade test form. Students were individually randomly assigned to one of the three test forms at each level. The fifth grade forms were administered before the seventh grade forms, in an effort to reduce the effect of anxiety on test performance. Two separate class periods were required to administer forms at the two levels. In addition, all students were given the Reading Subtest of the Metropolitan Achievement Tests. Fifth graders took Intermediate Form A and seventh graders Advanced Form A.

Results

Item analyses of each form were performed with responses

of students in the grade for which that form was constructed. These analyses indicated that sequence questions were often the most difficult items on the tests, especially for the fifth grade students. Discrimination and difficulty levels of most of the items measuring comprehension of main ideas and minor details and ability to draw inferences were, on the whole, in the acceptable range, and this fact was reflected in the KR20 reliabilities. Discrimination and difficulty levels for each form are presented in Corman (1973).

Table 2 presents the means and standard deviations of comprehension and rate on each final test form. KR20 reliability coefficients and correlations between comprehension and Metropolitan Test scores are also included in the table. Although the mean on Form 08 was higher, means of students on most of the forms developed for their grade level were close to 50% correct, indicating the possibility of a broad score range on the final forms. Standard deviations fell within a fairly narrow range.

Insert Table 2 about here

KR20 reliabilities of the six test forms ranged from .70 to .89 and all were significantly higher than .40 ($p < .01$). Validity coefficients of the three fifth grade forms ranged from .62 to .72 and were all significantly

TABLE 2

Means, Standard Deviations, Reliability and Validity Coefficients on Final Test Forms

	Comprehension		Words per minute				
Form	$\bar{X}$	SD	$\bar{X}$	SD	KR20	$r_{Met.}$	N
Grade 5							
05	52.05	19.45	179.29	61.62	.70	.62	54
06	56.75	19.19	196.09	71.06	.70	.72	56
08	65.33	21.55	187.98	53.85	.84	.64	48
Grade 7							
10	53.12	28.80	168.96	49.70	.89	.67	55
11	53.70	20.74	192.18	81.88	.76	.55	57
12	50.16	17.61	179.08	56.95	.70	.57	45

greater than .40 ($p < .05$). With regard to seventh grade forms, the validity coefficient for Form 10 was .67 ($p < .01$). The validity coefficients of Forms 11 and 12, however, were .55 and .57, respectively. Although these two coefficients were in the moderate range, they were not significantly higher than .40 for a sample this size. It is noteworthy that the validity coefficients for the two forms corresponding to 11 and 12 in the test revision phase (i.e., 7C and 7D) were .63 and .69, respectively ($p < .01$).

Analyses of variance on comprehension and rate tested the effects of Test Form, Difficulty Level (fifth versus seventh grade forms), and Content (fact versus fiction) for fifth and seventh graders separately. There was no significant difference on either comprehension or rate due to the particular test form a student received. This finding provided further evidence of equivalence among test forms. Fifth grade forms were found to be significantly easier than seventh grade forms in terms of comprehension but not rate, for students at both grade levels ($p < .001$). Fifth graders achieved significantly higher comprehension scores on the fictional selection, and the reading rate of seventh graders was higher on the fictional selection ($p < .01$).

Discussion

The primary goal of the study was achieved: three test forms for fifth and seventh grade students were

constructed, all of which showed evidence of reliability, content validity, a broad possible score range, and equivalence with the other forms developed for the same grade level.

The high degree of internal consistency of each form demonstrated the homogeneity of comprehension as measured by these tests. This finding is consistent with results of factor analytic studies of comprehension which have found comprehension, outside of word knowledge, to be a homogeneous trait (Davis, 1944, 1971). Further evidence of the homogeneity of comprehension on these tests was provided by high correlations among all four types of items obtained on scores with both fiction and factual material. This evidence of homogeneity confirms the decision to use total comprehension scores rather than subtest scores with these tests. Subtest scores on standardized reading tests have rarely been shown to possess a high degree of reliability (Smith & Dechant, 1961).

The construction of untimed measures of reading comprehension and rate, based on lengthy selections of two types of content and an uninterrupted three-minute time interval for reading, represents an important product of the study. None of the reading tests that are commonly used in schools provide this kind of measure of rate for students at these early grade levels. The statistical equivalence of the three test forms makes these forms useful

in research which tests reading achievement in the same students at two or three points in time.

Analysis of variance revealed the influence on comprehension of content of the reading material as well as its difficulty level. These findings suggest that those standardized reading tests, which measure reading skills on passages of only one difficulty level and of either fiction or factual content alone, may have limited generalizeability. Administration of tests at both grade levels to the same students enables the researcher to compare performance on material varying in difficulty. Similarly, inclusion of two types of content permits comparison of reading ability on fictional and factual material. Few standardized reading tests offer this opportunity.

References

Anderson, R. C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 2, 145-169.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971.

Buswell, G. T. Fundamental reading habits: A study of their development. Supplementary Educational Monographs, 1922, No. 21.

Chall, J. S. Interpretation of the results of standardized reading tests. In H. M. Robinson (Ed.), Supplementary Educational Monographs, 1958, No. 88

Chall, J. S. Learning to read: The great debate. New York: McGraw-Hill, 1967.

Clymer, T. What is 'reading'? Some current concepts. In H. M. Robinson (Ed.), The sixty-seventh yearbook of the National Society for the Study of Education. Part II. Chicago: University of Chicago Press, 1968.

Corman, L. Reading comprehension and rate: A study in measuring growth. Unpublished doctoral dissertation, Boston College, 1973.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. In W. A. Mehrens & R. L. Ebel (Eds.), Principles of educational and psychological measurement. Chicago: Rand McNally, 1967.

Davis. F. B. Fundamental factors of comprehension in reading. Psychometrika, 1944, 9, 185-197.

Davis, F. B. Psychometric research on comprehension in reading. In F. B. Davis (Ed.), The literature of research in reading with emphasis on models. New Brunswick, N.J.: Rutgers Graduate School of Education, 1971.

Furst, E. J. Constructing evaluation instruments. New York: Longmans, Green, 1958.

Guszak, F. T. Teachers' questions and levels of reading comprehension. In T. C. Barrett (Ed.), Perspectives in Reading, 1967, No. 8.

Harris, A. J. How to increase reading ability. New York: Longmans, Green, 1961.

Huey, E. B. The psychology and pedagogy of reading. Cambridge: M.I.T. Press, 1968. (Originally published: London: Macmillan, 1908).

Judd, C. H. Reading: Its nature and development. Supplementary Educational Monographs, 1918, 2 (4, Whole No. 10).

Robinson, F. P., & McCollum, F. H. Reading rate and comprehension accuracy as determinants of reading test scores. The Journal of Educational Psychology, 1934, 25, 154-157.

Simons, H. D. Reading comprehension: The need for a new perspective. Reading Research Quarterly, 1971, 6, 338-363.

Smith, H. P., & Dechant, E. V. Psychology in teaching reading. Englewood Cliffs, N.J.: Prentice-Hall, 1961

Spache, G. D. Toward better reading. Champaign, Ill.: Garrard, 1963.

Traxler, A. E. The nature and use of reading tests. Chicago: Science Research Associates, 1941. (a)

Traxler, A. E. Ten years of research in reading. New York: Educational Records Bureau, 1941. (b)

Traxler, A. E. Values and limitations of standardized reading tests. In H. M. Robinson (Ed.), Supplementary Educational Monographs, 1958, No. 88.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Footnotes

[1]This research was supported by grant OEG-0-8-080506-4597 from the Bureau of Education for the Handicapped, U.S. Office of Education and grant NE-G-00-3-0016 from the National Institute of Education, both under the Department of Health, Education, and Welfare.

[2]The items comprising main ideas included one item concerned with the main theme (often called the single main idea) and three other items concerned with major and significant elements in the story or article.